考 試 科 目	計算機概論	系 所 別	資訊管理學系/科技組	考 試 時 間	2 月 6 日（二）第二節

I. Multiple choice questions (40 points, 4 pts each)

1. The CAP theorem explains some of the competing requirements in a distributed system with replication. Which statements are accurate regarding this theorem?
   A) C means "Consistency", which guarantees that the nodes will have the same copies of a replicated data item in various transactions.
   B) A means "Atomicity", which guarantees that each transaction is a single unit that either succeeds completely or fails completely.
   C) P means "Partition tolerance", which guarantees that the system can continue operating even if the network fails and the nodes form disconnected partitions.
   D) Distributed databases like MongoDB and Cassandra tend to prioritize "C" and "P" at the cost of sacrificing "A".
   E) A MySQL database configured in the Master-Slave setting satisfies "C" and "A" but compromises "P".

   (1) AB　　(2) AC　　(3) CD　　(4) DE　　(5) BD

2. Which statements about SQL and NoSQL databases are most accurate?
   A) SQL is based on a structured query language with a fixed schema, while NoSQL is schema-less and can store unstructured data.
   B) NoSQL is designed to handle distributed data stores, making them a common choice for cloud storage and big data applications.
   C) Database normalization (i.e. 1NF, 2NF, 3NF, and 4NF) does not apply to NoSQL databases.
   D) NoSQL databases often sacrifice some of the ACID (Atomicity, Consistency, Isolation, Durability) properties for better scalability and performance.
   E) SQL is predominantly better at handling complex queries due to its rigid and well-defined schema.

   (1) ABD　　(2) BCD　　(3) ABC　　(4) ABDE　　(5) ABCDE

3. Which statements are most accurate regarding network protocol behavior and standards?
   A) OSPF is a routing protocol that calculates the shortest path for data packets to travel within an IP network using a path cost metric.
   B) ARP operates at the Internet layer to translate network addresses such as IP addresses into physical MAC (Media Access Control) addresses.
   C) ICMP is used for establishing and managing session states, often implemented at the Transport layer, alongside TCP and UDP.
   D) HTTP/2 introduces multiplexing of requests over a single TCP connection to reduce the amount of required connections.
   E) SSL/TLS protocols work at the Network layer to provide secure encryption capabilities for data packets

考 試 科 目	計算機概論	系 所 別	資訊管理學系/科技組	考 試 時 間	2 月 6 日(二)第二節

transmitted across networks.

(1) AB   (2) BC   (3) AD   (4) CD   (5) DE

4. What are the pros and cons of using threads compared to processes in parallel programming?
   A) Threads lead to more efficient system resource utilization than separate processes.
   B) Threads have lower context-switch overhead because they do not need to switch the memory space.
   C) Processes are more resilient to errors since one process crashing does not affect other processes. In contrast, thread errors can cause the entire process to crash.
   D) Threads need more complex synchronization mechanisms to manage shared resources and to avoid concurrency issues.
   E) Processes are generally easier to implement and debug, while threads can introduce complicated debugging scenarios because of shared state and asynchronous execution.

(1) ABE   (2) BCD   (3) ACE   (4) ABD   (5) ABCDE

5. Which statements about the Unified Modeling Language (UML) are most accurate?
   A) Sequence diagram shows the interaction between objects in a system by modeling messages exchanged between them over time.
   B) Activity diagram shows the workflow of a system by modeling activities, actions, and control flows.
   C) Use case diagram shows the functionality of a system by modeling actors, use cases, and their relationships.
   D) Composition represents a "has-a" relationship, while Aggregation represents an "is-part-of" relationship between the aggregated object and the aggregate object.
   E) A UML use case describes the overall behavior of the system from the perspective of the system.

(1) AED   (2) ABC   (3) BCD   (4) BDE   (5) ABCDE

6. Which statements are most accurate regarding cybersecurity?
   A) A rootkit is a set of specialized tools for system administrators to manage and monitor system security.
   B) A Zero-Day attack targets a vulnerability after the vulnerability is disclosed and before it is fixed.
   C) Post-quantum cryptography exploits quantum mechanics to secure communication and enhance cryptography.
   D) Social engineering is a type of attack that manipulates individuals into sharing confidential information that they should not share.
   E) Two-factor authentication (2FA) is a security process aiming to prevent the Man-in-the-Middle attack.

(1) AB   (2) BC   (3) CD   (4) BD   (5) AD

考 試 科 目	計算機概論	系 所 別	資訊管理學系/科技組	考 試 時 間	2 月 6 日（二）第二節

7. Which of the following statements accurately describes the mechanisms of the GPT (Generative Pre-trained Transformer) architecture?
   A) GPT uses an attention mechanism to focus on relevant parts of the input sequence and to facilitate context-aware text generation.
   B) GPT uses a pre-training phase on a large text corpus to learn general language patterns before fine-tuning on task-specific datasets.
   C) GPT uses recurrent neural layers that process input text sequentially, retaining state information across different positions in the text.
   D) GPT uses reinforcement learning to improve its predictions based on user feedback.
   E) GPT relies on stacked transformer blocks that utilize both self-attention and feed-forward neural networks within each block.

   (1) ABE   (2) BCD   (3) ACE   (4) BCE   (5) ABCDE

8. Which of the following statements are correct about Redundant Array of Independent Disks (RAID)?
   A) RAID 0 offers no data redundancy and distributes data across all disks to improve performance through stripping.
   B) RAID 1 substantially increases read and write speeds while offering complete data redundancy by mirroring data across all disks.
   C) RAID 5 requires at least three disks and provides data redundancy by distributing parity information across all disks in the array.
   D) RAID 6 is like RAID 5 but requires only two disks and uses double striping instead of parity for fault tolerance.
   E) RAID 10 (1+0) can tolerate the failure of at least one drive in each mirrored pair and typically requires more storage capacity than RAID 5 or 6.

   (1) ABE   (2) BCD   (3) ACE   (4) BCE   (5) ABCDE

9. Which statements about software development are most accurate?
   A) Waterfall development sets clear milestones in each stage and moves to the next stage only after the goal of the current stage is fulfilled.
   B) Waterfall development appreciates a high degree of customer involvement so that the project outcome can be adjusted in a timely manner.
   C) Waterfall development is more suitable than Agile when multiple software components must be designed in parallel for final integration.
   D) Agile allows the team members to be involved with other work depending on the phases, while Waterfall demands highly devoted team members throughout the development.

考 試 科 目	計算機概論	系 所 別	資訊管理學系/科技組	考 試 時 間	2 月 6 日（二）第二節

E) Agile relies on careful documentation and testing to ensure the quality and understanding of the software deliverables.

(1) AD   (2) BC   (3) AE   (4) AC   (5) CE

10. What statements about open-source licenses are most accurate?
   A) If the software uses GPL-licensed code, it must disclose source code when it is distributed.
   B) The MIT license requires including the copyright and permission notice in all copies of the software.
   C) The Apache license allows the software to be patented but prohibits using the licensor's trademark.
   D) GPL allows both the right to patent the software and the right to use the licensor's trademark.
   E) MIT and Apache allow commercial use of the licensed software, but GPL does not.

(1) ABDE   (2) ABCE   (3) ABCD   (4) BCDE   (5) ABCDE

II. Short-Answer Questions (60 points)

1. (5 points) What problems would imbalanced training data cause in machine learning? How do you mitigate these problems? Give at least one problem and one solution.

2. (10 points) Explain the major differences between virtual machines (VMs) and containers. Give a scenario where VMs are preferable to containers, and a scenario where containers are preferable to VMs.

3. (12 points) Give a short definition and at least one example for each of the following:
   - Data as a Service (DaaS)
   - Software as a Service (SaaS)
   - Platform as a Service (PaaS)
   - Infrastructure as a Service (IaaS)

4. (15 points) Inspect the following multi-threaded Java program and answer the questions.

考 試 科 目	計算機概論	系 所 別	資訊管理學系/科技組	考 試 時 間	2月 6 日（二）第二節

```java
public class Main {
  public static void main(String[] args) {
    Lock lock = new Lock();
    for (int i = 1; i <= 10; i++) {
      final int tid = i;
      Thread t = new Thread(new Runnable() {
        public void run() {
          lock.acquire();
          System.out.println("Thread " + tid);
          lock.release();
        }
      });
      t.start();
    }
  }
}
```

```java
class Lock {
  Object lock = new Object();
  int now = 0, next = 0;
  void acquire() {
    int id;
    synchronized(this) {
      id = next;
      next++;
    }
    while(id != now) {}
  }
  void release() {
    now++;
  }
}
```

4.1 (5 pts) Argue that no two threads can print to the terminal at the same time.

4.2 (5 pts) Is the program still thread-safe after we change `synchronized(this)` to `synchronized(lock)`? Why?

4.3 (5 pts) Is it true that every thread eventually prints a message? Why?

5.

(4 points) Consider a table `Customers` that contains fields `CustomerID` and `Country`. Write a MySQL code to list the number of customers in each country, ordered by the country with the most customers first.

Sample output:

```
Country | Count
----------------
Taiwan  | 1236
USA     | 452
Japan   | 87
Korea   | 31
```

6.

(4 points) Suppose that you have two tables, `Customers` and `Orders`. Each entry of `Customers` contains two fields: `CustomerID` and `CustomerName`. Each entry of `Orders` contains two fields: `CustomerID` and an `OrderID`. A customer can be related to several orders through `CustomerID`. Write a MySQL code to list the customer names and the number of orders associated with each customer.
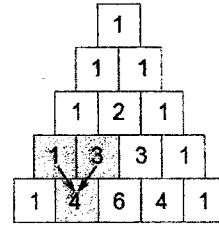
Sample output:

```
Name  | Count
--------------
Alice | 120
Bob   | 75
Carl  | 256
```

7. (10 points) Pascal's triangle is defined as below. The row number starts from 1 (e.g., Row 1 is [1], Row 2 is [1 1], Row 3 is [1 2 1], etc.). Write a function that takes a row number $k \geq 1$ and returns Row k of the triangle. Your program should run in $O(k^2)$ time and $O(k)$ space.

考　試　科　目	計算機概論	系 所 別	資訊管理學系/科技組	考 試 時 間	2月6日(二)第二節

Examples:

   Input: 1   Output: [1]

   Input: 3   Output: [1 2 1]

   Input: 4   Output: [1 3 3 1]

```
        1
      1   1
    1   2   1
  1   3   3   1
1   4   6   4   1
```